

ON EXACT MAXIMUM-LIKELIHOOD DETECTION FOR NON-COHERENT MIMO WIRELESS SYSTEMS: A BRANCH-ESTIMATE-BOUND OPTIMIZATION FRAMEWORK

Weiyu Xu, Mihailo Stojnic and Babak Hassibi

Electrical Engineering Department, California Institute of Technology

ABSTRACT

Fast fading wireless environments pose a great challenge for achieving high spectral efficiency in next generation wireless systems. Joint maximum-likelihood (ML) channel estimation and signal detection is of great theoretical and practical interest, especially for multiple-input multiple-output (MIMO) systems where the multiple channel coefficients need to be estimated. However, this is a hard combinatorial optimization problem, for which obtaining efficient *exact* algorithms has been elusive for the general MIMO systems. In this paper, we propose an efficient branch-estimate-bound non-coherent optimization framework which *provably* achieves the exact ML joint channel estimation and data detection for general MIMO systems. Numerical results indicate that the exact joint ML method can achieve substantial performance improvements over suboptimal methods including iterative channel estimation and signal detection. We also derive analytical bounds on the computational complexity of the new exact joint ML method and show that its average complexity approaches a constant times the length of the coherence time, as the SNR approaches infinity.

1. INTRODUCTION

The goal of achieving high-speed reliable data transmission over highly dynamic wireless media has generated a lot of research activities in the information theory, communications and signal processing communities. One of the major challenges there is due to fading, where the wireless channels may vary over time. Usually the time-varying channels are estimated from the training sequences and then used for signal detection, but relying solely on the training sequences may result in inaccurate channel estimation, especially in fast fading scenarios. Occasionally, the channels may even change so rapidly that training and accurate channel tracking become infeasible. In such cases, joint channel estimation and data detection can enhance the system performance considerably [6, 7, 8, 10, 11, 12] and potentially reduces the resources dedicated to training.

MIMO systems have received a significant amount of attention in recent years for the promise of greatly increasing spectral efficiency and exploiting both transmit and receive diversity when the channel state information is known. However, the wireless channel uncertainties, especially in fast fading scenarios, pose challenges in achieving the benefits in the MIMO systems [1][2] [5][16]. It usually takes more resources like power and bandwidth to deal with the unknown channel coefficients in multiple antenna systems. To boost the spectral efficiency and enhance the system performance, it is very useful to consider the problem of joint ML channel estimation and data detection for MIMO systems.

This work was supported in part by the National Science Foundation under grant no. CCR-0729203, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking.

Significant progress has recently been made in this direction. For example, suboptimal iterative receiver structures were considered in [6, 13, 14]. In [7], the joint *exact* ML channel estimation and data detection problem was shown to be an integer least-squares problem for single-input multiple-output (SIMO) systems under constant modulus constellations, where the sphere decoder was proposed for finding the joint ML solution. In [10], the sphere decoder and a semidefinite relaxation approach (SDR) were used for blind detection of orthogonal space-time block codes (OSTBC) by using the orthogonal structure of the OSTBC. In [9], the authors showed that the joint ML estimation and detection problem admits certain polynomial-time algorithm in the coherence time for constant energy constellations in the single transmit antenna wireless fading systems.

Despite all these developments, obtaining efficient *exact* joint ML detection algorithms have remained elusive for general MIMO systems. In the suboptimal iterative estimation and detection framework, the solutions are not guaranteed to be the joint ML solution and not much is known about the gap between these iterative schemes and the exact joint ML scheme. In fact, as will be seen below, the exact joint ML detection problem turns out to be a complicated combinatorial optimization problem and generally can not be transformed into a standard integer least-squares problem as in [7] [10], which makes sphere decoders inapplicable to general non-coherent MIMO systems. This is in sharp contrast to the coherent MIMO cases where the problem is exactly an integer least-squares problem [3][4]. To the best of our knowledge, there are no working algorithms with numerical results reported in the literature for the high-dimensional exact joint ML detection problem of *general* MIMO systems.

In this paper, we propose an efficient branch-estimate-bound non-coherent optimization framework which provably achieves the exact ML joint channel estimation and data detection in MIMO systems. Analysis shows that for any data length, if the SNR is high enough, the average complexity of the algorithm approaches a constant times the data length while achieving the optimal performance. Simulation results verify the efficacy of the new algorithm.

2. PROBLEM FORMULATION

Let us consider a MIMO system with M transmit antennas and N receive antennas. Let T be the length of a data packet during which the channel remains constant. Then the channel output is written as

$$\mathbf{X} = \mathbf{H}\mathbf{S}^* + \mathbf{W}, \quad (1)$$

where $\mathbf{H} \in \mathcal{C}^{N \times M}$ is the MIMO channel matrix, $\mathbf{S}^* \in \mathcal{C}^{M \times T}$ is the transmitted symbol matrix, and $\mathbf{W} \in \mathcal{C}^{N \times T}$ is an additive noise matrix whose elements are assumed to be i.i.d. complex Gaussian random variables. We also assume that except for the possible

training symbols, the entries of \mathbf{S}^* are i.i.d. symbols from a certain constellation Ω (such as QPSK or 16-QAM). We should remark, however, that the results in this paper can be extended to cases where \mathbf{S}^* is taken from some general linear dispersion space-time codes.

We assume a deterministic unknown channel model for \mathbf{H} where \mathbf{H} is assumed to be a deterministic unknown and no priori information about \mathbf{H} is known [6][10] [15]. This deterministic unknown channel model is preferred in some scenarios over the assumption of stochastic channel gains [6][10] [15]. Wireless channels may obey different distributions, like Rayleigh fading, Rician fading, Nakagami fading or other fading statistics and there possibly are correlations between the channel coefficients across the N receive antennas. These statistics are needed in the joint ML detection for a stochastic channel model, but they are not always available to the receiver. In fact, if we assume that \mathbf{H} is complex Gaussian distributed, the joint ML channel estimation and data detection problem can still be transformed to a special case for which the methods of this paper can be applied. However, for the simplicity of presentation, we will stick with the unknown deterministic model for \mathbf{H} .

Under these assumptions, the problem of joint ML channel estimation and data detection for MIMO channels reduces to solving the following optimization problem

$$\min_{\mathbf{H}, \mathbf{S}^* \in \Omega^{M \times T}} \|\mathbf{X} - \mathbf{H}\mathbf{S}^*\|^2, \quad (2)$$

where $\Omega^{M \times T}$ denotes the $M \times T$ -dimensional signal lattice. In fact for any given transmitted symbol \mathbf{S}^* , the channel vector \mathbf{H} that minimizes (2) is given by the least-squares solution,

$$\hat{\mathbf{H}} = \mathbf{X}(\mathbf{S}^*)^\dagger, \quad (3)$$

where $(\cdot)^\dagger$ denotes the Moore-Penrose pseudoinverse of a matrix. Using the relation

$$(\mathbf{S}^*)^\dagger = \mathbf{S}(\mathbf{S}^*\mathbf{S})^\dagger, \quad (4)$$

We have

$$\hat{\mathbf{H}} = \mathbf{X}\mathbf{S}(\mathbf{S}^*\mathbf{S})^\dagger. \quad (5)$$

Substituting (5) into (2), we get

$$\min_{\mathbf{H}, \mathbf{S}^*} \|\mathbf{X} - \mathbf{H}\mathbf{S}^*\|^2 = \min_{\mathbf{S}^*} \|\mathbf{X}(\mathbf{I} - \mathbf{S}(\mathbf{S}^*\mathbf{S})^\dagger\mathbf{S}^*)\|^2 \quad (6)$$

$$= \min_{\mathbf{S}^*} \text{tr}(\mathbf{X}(\mathbf{I} - \mathbf{S}(\mathbf{S}^*\mathbf{S})^\dagger\mathbf{S}^*)\mathbf{X}^*) \quad (7)$$

$$= \text{tr}(\mathbf{X}\mathbf{X}^*) - \max_{\mathbf{S}^*} \text{tr}((\mathbf{S}^*\mathbf{S})^\dagger\mathbf{S}^*\mathbf{X}^*\mathbf{X}\mathbf{S}), \quad (8)$$

where $\text{tr}(\cdot)$ is the trace of a matrix. As we can see from (8), the joint ML channel estimation and data detection problem is a combinatorial problem involving $|\Omega|^{MT}$ hypothesis tests and it appears impossible to exactly solve it efficiently for large Ω and MT . If we consider the SIMO or OSTBC systems with constant modulus modulation, by the orthogonal structure in the rows of \mathbf{S}^* , the minimization of (8) over \mathbf{S}^* can be simplified to solving [7, 10]

$$\max_{\mathbf{S} \in \Omega^{M \times T}} \text{tr}(\mathbf{S}^*\mathbf{X}^*\mathbf{X}\mathbf{S}), \quad (9)$$

If ρ is larger than the maximum eigenvalue of $\mathbf{X}^*\mathbf{X}$, (9) is equivalent to the following integer least-squares problem (for constant modulus constellations)

$$\min_{\mathbf{S} \in \Omega^{M \times T}} \text{tr}(\mathbf{S}^*(\rho\mathbf{I} - \mathbf{X}^*\mathbf{X})\mathbf{S}), \quad (10)$$

After a Cholesky decomposition, the sphere decoding algorithm can solve the integer least-squares much more efficiently than exhaustive search. The readers are referred to [7] for details of sphere decoding procedures. However, with the presence of the term $(\mathbf{S}^*\mathbf{S})^\dagger$, the minimization of (8) can not be reduced to the problem (9), or (10) in general MIMO systems.

3. THE BRANCH-ESTIMATE-BOUND OPTIMIZATION FRAMEWORK

In this paper, instead, we start by considering the original problem given in (2). For \mathbf{X} , \mathbf{S}^* and \mathbf{W} , we denote the parts corresponding to their first i ($1 \leq i \leq T$) time indices (columns) as $\mathbf{X}_{(i)}$, $\mathbf{S}_{(i)}^*$ and $\mathbf{W}_{(i)}$ while \mathbf{X}_i , \mathbf{S}_i^* and \mathbf{W}_i will be reserved for their part corresponding to exactly the i -th time index (column). Let us consider a partial data sequence $\mathbf{S}_{(i)}^*$ up to time index i and define $M_{\mathbf{S}_{(i)}^*}$ as

$$M_{\mathbf{S}_{(i)}^*} = \text{tr}(\mathbf{X}_{(i)}\mathbf{X}_{(i)}^*) - \text{tr}((\mathbf{S}_{(i)}^*\mathbf{S}_{(i)}^*)^\dagger\mathbf{S}_{(i)}^*\mathbf{X}_{(i)}\mathbf{X}_{(i)}\mathbf{S}_{(i)}^*), \quad (11)$$

namely the partial joint ML metric up to the time index i for \mathbf{S}^* . Clearly, the solution minimizing this partial joint ML metric is not globally optimal. But we have the following lemma:

Lemma 1 *Let R be the optimal value for the optimization problem in (2). If $M_{\mathbf{S}_{(i)}^*} > R$, then $\mathbf{S}_{(i)}^*$ can not be the first i signal vectors of the ML solution $\hat{\mathbf{S}}^*$ to (2).*

Proof: Suppose that $\mathbf{S}_{(i)}^*$ are the first i symbols of the ML solution $\hat{\mathbf{S}}^*$ and denote the optimal channel gains corresponding to $\hat{\mathbf{S}}^*$ as $\hat{\mathbf{H}}$. Then

$$\begin{aligned} R &= \|\mathbf{X}_{(i)} - \hat{\mathbf{H}}\hat{\mathbf{S}}_{(i)}^*\|^2 + \sum_{j=i+1}^T \|\mathbf{X}_j - \hat{\mathbf{H}}\hat{\mathbf{S}}_j^*\|^2 \\ &\geq \min_{\mathbf{H}} \|\mathbf{X}_{(i)} - \mathbf{H}\hat{\mathbf{S}}_{(i)}^*\|^2 + \sum_{j=i+1}^T \|\mathbf{X}_j - \hat{\mathbf{H}}\hat{\mathbf{S}}_j^*\|^2 \\ &\geq \min_{\mathbf{H}} \|\mathbf{X}_{(i)} - \mathbf{H}\hat{\mathbf{S}}_{(i)}^*\|^2 = M_{\hat{\mathbf{S}}_{(i)}^*} = M_{\mathbf{S}_{(i)}^*}, \end{aligned} \quad (12)$$

where the last two equalities are due to (8) and the assumption $\hat{\mathbf{S}}_{(i)}^* = \mathbf{S}_{(i)}^*$. But since $M_{\mathbf{S}_{(i)}^*}$ is larger than R , we have a contradiction. ■

From Lemma 1, if the value R of the optimization problem (2) can be estimated, the search of the blind ML solution $\hat{\mathbf{S}}^*$ will be restricted to the offsprings of those partial sequences $\mathbf{S}_{(i)}^*$ which satisfy $M_{\mathbf{S}_{(i)}^*} \leq R$. This motivates a trimmed search over a signal tree of T layers, where each tree node at the i -th layer corresponds to a specific partial sequence $\mathbf{S}_{(i)}^*$ and each tree node at the intermediate layer has $|\Omega|^M$ offsprings to the next layer. From Lemma 1, (11) is then a lower bound (based on the locally optimal channel estimation for $\mathbf{S}_{(i)}^*$) for the global metric (2). We can then trim the offsprings of $\mathbf{S}_{(i)}^*$ from consideration if (11) is larger than R . We call this idea the branch-estimate-bound method. Before introducing our algorithm, we first number the $|\Omega|^M$ combinations of the constellation points from the M transmitter antennas by $1, 2, \dots, |\Omega|^M$ and treat them as a big constellation set Ψ , where the k -th ($1 \leq k \leq |\Omega|^M$) vector constellation point is denoted by $\Psi(k)$. We perform a depth-first search on the signal tree for the joint ML solution.

Branch-Estimate-and-Bound Algorithm

Input: radius r , \mathbf{X} , constellation Ω and a $1 \times T$ index vector I

1. Set $i = 1$, $I(i) = 1$ and set $\mathbf{S}_i^* = \Psi(I(i))$.
2. (Computing the lower bound based on local estimation) Compute the metric $M_{\mathbf{S}_{(i)}^*}$ based on (11). If $M_{\mathbf{S}_{(i)}^*} > r$, go to the backtracking Step 3 and search over other branches; else, go to 4;
3. (Backtracking) Find the largest $1 \leq j \leq i$ such that $I(j) < |\Omega|^M$. If there exists such j , set $i = j$ and go to 5; else go to 6.

4. If $i = T$, store current $\mathbf{S}_{(T)}^*$, update $r = M_{\mathbf{S}_{(T)}^*}$ and go to 3; else set $i = i + 1$, $I(i) = 1$ and $\mathbf{S}_i^* = \Psi(I(i))$, go to 2.
5. Set $I(i) = I(i) + 1$ and $\mathbf{S}_i^* = \Psi(I(i))$. Go to 2.
6. If any sequence $\mathbf{S}_{(T)}^*$ is ever found in 4, output the latest stored full-length sequence as the ML solution; otherwise, double r and go to 1.

3.1. Choice of the initial radius r

The structure of the new joint ML channel estimation and data detection algorithm easily suggests a probabilistic choice of the search radius. We know that $\|\mathbf{W}\|^2$ is chi-square distributed with $2NT$ degrees of freedom, it is natural to choose the radius r such that $P(\|\mathbf{W}\|^2 > r^2) \leq 1 - \epsilon$. Since the set (\mathbf{H}, \mathbf{S}) is a feasible point for the optimization problem (2), the objective value R of the solution to the optimization problem in (2) is sure to be no bigger than $\|\mathbf{W}\|^2$, we will guarantee finding the blind ML data sequence with probability at least $1 - \epsilon$ under the initial radius r .

3.2. Constant update complexity per tree node w.r.t. T

As we can see, the main source of complexity comes from Step 2 where the metric $M_{\mathbf{S}_{(i)}^*}$ is computed, which at first sight needs computational efforts $O(T)$. But actually the complexity for Step 2 can be made independent of T . To see this, we can efficiently update $M_{\mathbf{S}_{(i)}^*}$ at each tree node with constant computational costs with respect to T by propagating the matrix \mathbf{A}_i , which is a $N \times M$ matrix defined as $\mathbf{A}_i = \sum_{j=1}^i \mathbf{X}_j \mathbf{S}_j$, and the $M \times M$ matrix \mathbf{B}_i , which is defined as $\mathbf{B}_i = \sum_{j=1}^i \mathbf{S}_j^* \mathbf{S}_j$. Then we can update \mathbf{A}_{i+1} sequentially as $\mathbf{A}_{i+1} = \mathbf{A}_i + \mathbf{X}_{i+1} \mathbf{S}_{i+1}$, \mathbf{B}_{i+1} sequentially as $\mathbf{B}_{i+1} = \mathbf{B}_i + \mathbf{S}_{i+1}^* \mathbf{S}_{i+1}$. The metric $M_{\mathbf{S}_{(i+1)}^*}$ is computed as $\text{tr}(\mathbf{X}_{(i+1)} \mathbf{X}_{(i+1)}^*) - \text{tr}(\mathbf{B}_{i+1} \mathbf{A}_{i+1}^* \mathbf{A}_{i+1})$. Notice that $\text{tr}(\mathbf{X}_{(i)} \mathbf{X}_{(i)}^*)$, namely the energy of $\mathbf{X}_{(i)}$, can be sequentially pre-computed at the beginning and used in the whole search. So the computational cost of updating each tree node's metric does not depend on the search dimension T . Also, by the matrix inversion lemma, the update of $\mathbf{B}_{i+1}^\dagger = (\mathbf{B}_i + \mathbf{S}_{i+1}^* \mathbf{S}_{i+1})^\dagger$ can be efficiently done with complexity $O(M^2)$ when \mathbf{B}_i is of rank M , which is usually the case when there is a rank- M training symbol matrix.

When we search along the tree in the Branch-Estimate-and-Bound algorithm, the corresponding matrix \mathbf{A}_i and \mathbf{B}_i for each tree node on the current path is stored. So when we do backtracking on that path, we do not need to recompute \mathbf{A}_{i-1} and \mathbf{B}_{i-1} in Step 2 for computing $M_{\mathbf{S}_i^*}$.

4. COMPUTATIONAL COMPLEXITY

In this section we provide an upper bound on the expected value of the number of arithmetic operations required by our algorithm. Given that the number of the arithmetic operations performed per each node of the search tree is roughly upper-bounded by a constant dependent on N and M , our problem simplifies to finding the expected value of the number of the nodes inside the search radius r . We will solve this problem by finding the expected value of the number of nodes per each level in the search radius. To simplify analysis, we assume i.i.d. complex Gaussian channel gains as $\sqrt{\frac{\rho}{Mc}} \bar{\mathbf{H}}$, where ρ is the SNR, $\bar{\mathbf{H}}$ is a matrix with i.i.d. complex Gaussian random variables of unit variance, c is the average energy of each component of the transmitted matrix \mathbf{S}^* and M is the number of transmit antenna. We remark that although the deterministic channel

model was used in the receiver design because the receiver does not know the channel statistics, we need to assume some distribution for the channel coefficients in the performance and complexity analysis. Let the expected number of the nodes at i -th level of the search tree be C_i . Recall that every node in the tree has a sequence of specific vector constellation points from Ψ which lead to that node. Let $\bar{\mathbf{S}}_{(i)}$ ($\bar{\mathbf{S}}_{(i)} \neq \mathbf{S}_{(i)}^*$) be the sequence of vector constellation points at a node of the i -th level which has the largest probability of being inside the search radius. Let P_i be the probability that this node is in the tree. Then it is easy to see that

$$C_i \leq 1 + (q^i - 1)P_i \quad (13)$$

where q denotes the alphabet set size in our vector constellation. From (13) it clearly follows that establishing an upper bound on P_i is sufficient to establish an upper bound on C_i . Hence, we will focus on bounding P_i .

From the description of our algorithm it follows that

$$P_i = \Pr(\|\mathbf{X}_{(i)} (I - \bar{\mathbf{S}}_{(i)} (\bar{\mathbf{S}}_{(i)}^* \bar{\mathbf{S}}_{(i)})^\dagger \bar{\mathbf{S}}_{(i)}^*)\|_F^2 \leq r^2)$$

where $\mathbf{X}_{(i)}$ denotes the first i columns of the matrix \mathbf{X} and r^2 is the squared value of the search radius. We can further write

$$P_i = \Pr\left(\text{Tr}\left(\mathbf{X}_{(i)} (I - \bar{\mathbf{S}}_{(i)} (\bar{\mathbf{S}}_{(i)}^* \bar{\mathbf{S}}_{(i)})^\dagger \bar{\mathbf{S}}_{(i)}^*) \mathbf{X}_{(i)}^*\right) \leq r^2\right). \quad (14)$$

Recall that

$$\mathbf{X}_{(i)} = \sqrt{\frac{\rho}{Mc}} \bar{\mathbf{H}} \mathbf{S}_{(i)}^* + \mathbf{W}_{(i)}. \quad (15)$$

Combining (14) and (15) we have

$$P_i = \Pr\left(\text{Tr}\left(\left[\begin{array}{c} \bar{\mathbf{H}}^* \\ \mathbf{W}_{(i)}^* \end{array}\right]^* Q_{(i)} \left[\begin{array}{c} \bar{\mathbf{H}}^* \\ \mathbf{W}_{(i)}^* \end{array}\right]\right) \leq r^2\right) \quad (16)$$

where

$$Q_{(i)} = \left[\sqrt{\frac{\rho}{Mc}} \mathbf{S}_{(i)}^* \right] \left(I - \bar{\mathbf{S}}_{(i)} (\bar{\mathbf{S}}_{(i)}^* \bar{\mathbf{S}}_{(i)})^\dagger \bar{\mathbf{S}}_{(i)}^*\right) \left[\sqrt{\frac{\rho}{Mc}} \mathbf{S}_{(i)} \quad I\right].$$

Using the Chernoff bound the right hand side of (16) can be bounded in the following way

$$\begin{aligned} P_i &\leq \min_{\mu \geq 0} e^{\mu r^2} E e^{-\mu \text{Tr}\left(\left[\begin{array}{c} \bar{\mathbf{H}}^* \\ \mathbf{W}_{(i)}^* \end{array}\right]^* Q_{(i)} \left[\begin{array}{c} \bar{\mathbf{H}}^* \\ \mathbf{W}_{(i)}^* \end{array}\right]\right)} \\ &= \min_{\mu \geq 0} \int e^{-\mu \text{Tr}\left(\left[\begin{array}{c} \bar{\mathbf{H}}^* \\ \mathbf{W}_{(i)}^* \end{array}\right]^* (I + \mu Q_{(i)}) \left[\begin{array}{c} \bar{\mathbf{H}}^* \\ \mathbf{W}_{(i)}^* \end{array}\right]\right)} \frac{d\bar{\mathbf{H}}^* d\mathbf{W}_{(i)}^*}{e^{-\mu r^2} \pi^{N(i+M)}} \\ &= \min_{\mu \geq 0} \frac{e^{\mu r^2}}{\det(I + \mu Q_{(i)})^N}. \end{aligned} \quad (17)$$

For the determinant in the denominator, we have

$$\begin{aligned} \det(I + \mu Q_{(i)}) &= \det\left(I + \mu \left(I + \frac{\rho}{Mc} \mathbf{S}_{(i)} \mathbf{S}_{(i)}^*\right) (I - \bar{\mathbf{S}}_{(i)} (\bar{\mathbf{S}}_{(i)}^* \bar{\mathbf{S}}_{(i)})^\dagger \bar{\mathbf{S}}_{(i)}^*)\right) \\ &= \det\left(I + \mu \left(I + \frac{\rho}{Mc} \mathbf{S}_{(i)} \mathbf{S}_{(i)}^*\right) (I - \bar{\mathbf{S}}_{(i)} (\bar{\mathbf{S}}_{(i)}^* \bar{\mathbf{S}}_{(i)})^\dagger \bar{\mathbf{S}}_{(i)}^*)\right) \end{aligned} \quad (18)$$

Combining (13), (17), and (18) we have an upper bound on the expected number of nodes inside the search radius, namely one plus

$$\min_{\mu \geq 0} \frac{(q^i - 1) e^{\mu r^2}}{(\det(I + \mu (I + \frac{\rho}{Mc} \mathbf{S}_{(i)} \mathbf{S}_{(i)}^*) (I - \bar{\mathbf{S}}_{(i)} (\bar{\mathbf{S}}_{(i)}^* \bar{\mathbf{S}}_{(i)})^\dagger \bar{\mathbf{S}}_{(i)}^*)))^N}.$$

Setting $\mu = \frac{N_i}{r^2}$ and assuming $\mathbf{S}_{(i)}\mathbf{S}_{(i)}^*(\mathbf{I} - \bar{\mathbf{S}}_{(i)}(\bar{\mathbf{S}}_{(i)}^*\bar{\mathbf{S}}_{(i)})^\dagger\bar{\mathbf{S}}_{(i)}^*) \neq \mathbf{0}$, it is not hard to verify that

$$\lim_{\rho \rightarrow \infty} C_i \leq 1. \quad (19)$$

This effectively means that for high enough SNR, only nodes corresponding to the sequence \mathbf{S}^* will be inside the search radius. Since we perform a constant number of arithmetic operations w.r.t. T per tree node, the average overall number of arithmetic operations will be approaching a constant times T when the SNR approaches infinity, which is an interesting counterpart to the behavior of tree-searching sphere decoder for coherent MIMO channels in the high SNR region.

5. SIMULATION RESULTS

In this section, we give simulation results for the performance and complexity of the new blind ML decoder. First, We investigate its performance and complexity in MIMO systems employing QPSK where we assume $M = 2$, $N = 6$ and $T = 12$. In the simulations, the channel coefficients are generated as i.i.d complex Gaussian random variables. To resolve the channel ambiguity, we embed two columns of known pilot symbols in the data block (otherwise, even the exact ML solution will be in very bad error region) and they are chosen to be orthogonal optimum training symbols as discussed in [2]. Figure 1 shows the error performance comparison between the branch-estimate-and-bound algorithm, the iterative scheme and the coherent detection case. In the simulation for the iterative joint channel estimation and data detection scheme [14], the least-squares (LS) channel estimation was first initialized from the training blocks based on (3) and then the so-obtained channel estimation was used for coherent data detection. After that, the channel is re-estimated by further using the detection decisions. The estimation-detection procedure was iterated for a preimposed 20 times in the simulations. The exact joint ML scheme performs about 1.25dB better than the iterative scheme at the BER 10^{-3} for this QPSK MIMO system.

In Figure 2, we give the BER performance of our algorithm for a 16-QAM MIMO system with $M = 2$, $N = 6$ and $T = 7$. Again, two columns of known pilots are embedded in each data block. We can see that the new algorithm works closer to the known channel case, while having obvious performance advantage over the iterative joint channel estimation and data detection [14], about 3dB better at the BER 10^{-3} . Notice for each block, we need to test $16^{2 \times 5} \approx 10^{12}$ hypothesis assumptions in exhaustive search for solving the joint ML problem (2), which is almost infeasible for getting this result. In Figure 3, we show the average number of floating-point operations (FLOPS) in decoding one data block by using the branch-estimate-and-bound algorithm. Compared with exhaustive search, the only other known exact ML method for this system in the literature, the branch-estimate-and-bound algorithm is more efficient by several orders of magnitude. In the high SNR region, the complexity of the branch-estimate-and-bound algorithm is comparable with the complexity of suboptimal iterative channel estimation and data detection schemes while in the low SNR region, the branch-estimate-and-bound exact ML algorithm will have higher computational complexity than the iterative method.

6. SUMMARY AND DISCUSSION

In this paper, we investigate the problem of joint ML channel estimation and data detection for MIMO wireless and propose a new efficient branch-estimate-and-bound optimization framework which

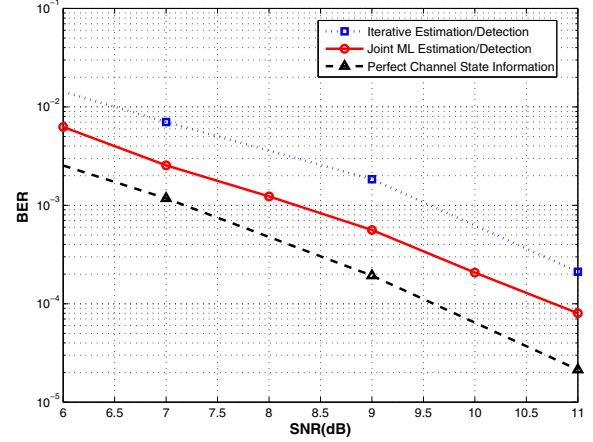


Fig. 1. Joint ML Channel Estimation and Data Detection Performance for QPSK MIMO Systems with $M = 2$, $N = 6$ and $T = 12$

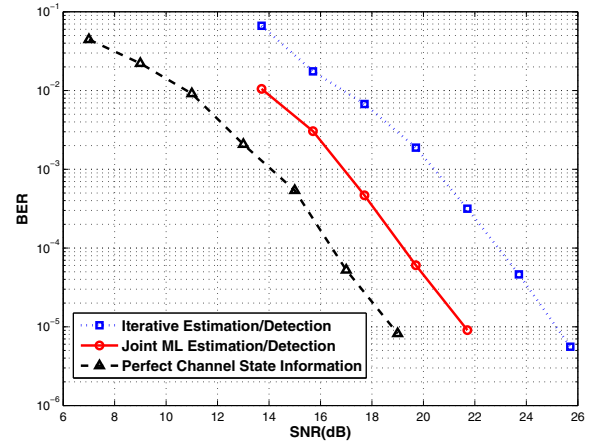


Fig. 2. Joint ML Channel Estimation and Data Detection Performance for 16-QAM MIMO Systems with $M = 2$, $N = 6$ and $T = 7$

achieves the exact joint ML performance for general MIMO systems. Numerical results verified that our proposed algorithm solves the exact joint ML channel estimation and data detection problem for noncoherent MIMO systems efficiently. It is shown analytically that as the SNR grows, the average complexity of the exact joint ML channel estimation and data detection algorithm approaches a constant times the data length T . Although the branch-estimate-and-bound framework is efficient with respect to T , we have $|\Omega|^M$ branches from each intermediate node, which can still be a large number if $M > 4$ or $|\Omega| \geq 16$. It is thus interesting to explore how to bring this branching number down in this framework. The results in this paper can also be extended in different ways, for example, to stochastic channel models or to more efficient algorithms for specific space-time codes.

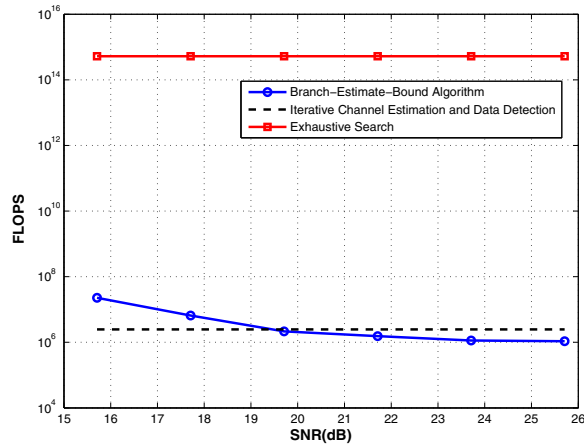


Fig. 3. FLOPS Count for the Joint Channel Estimation and Data Detection Algorithms for 16-QAM MIMO Systems with $M = 2$, $N = 6$ and $T = 7$

7. REFERENCES

- [1] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543-564, Mar. 2000.
- [2] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Communications, Control, and Computing*, Sept. 22-24, 1999.
- [3] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2389-2402, Oct. 2003.
- [4] A. D. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 933-953, Mar. 2006.
- [5] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple- antenna wireless links?," *IEEE. Trans. Inform. Theory*, vol. 49, pp. 951-963, Apr. 2003.
- [6] P. Stoica and G. Ganesan, "Space-time block codes: Trained, blind, and semi-blind detection," *Digital Signal Process.*, vol. 13, pp. 93-105, 2003.
- [7] H. Vikalo, B. Hassibi and P. Stoica, "Efficient joint maximum-likelihood channel estimation and signal detection," *IEEE Transactions on Wireless Communications*, vol 5, no 7, pp. 1838-1845, Jul 2006.
- [8] P.Stoica, H.Vikalo and B.Hassibi,"Joint maximum-likelihood channel estimation and signal detection for SIMO channels," in *Proceedings of 2003 International Conference on Acoustics,speech and signal processing*, vol.4,2003,pp. 13-16.
- [9] I.Motedayen,A.Krishnamoorthy and A.Anastasopoulos, "Optimal Joint Detection/Estimation in Fading Channels with Polynomial Complexity," *IEEE Transactions on Information Theory*,vol. 53, no. 1, pp. 209-223 January 2007;
- [10] W.-K. Ma, B.-N. Vo, T.N. Davidson, and P.C. Ching, "Blind ML detection of orthogonal space-time block codes: High-performance, efficient implementations," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 738-751, Feb. 2006.
- [11] A. L.Swindlehurst and G. Leus, "Blind and semi-blind equalization for generalized space-time block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2589-2498, 2002.
- [12] S. Shahbazpanahi, A. B. Gershman, and J. H. Manton, "Closed-form blind MIMO channel estimation for orthogonal space-time block codes," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4506-4517, Dec. 2005.
- [13] C.Cozzo and B.L.Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. on Communications*, vol.51, no.8, pp.1266-1270, Aug.2003.
- [14] E.G.Larsson, P.Stoica and J.Li,"Orthogonal space-time block codes:Maximum likelihood detection for unknown channels and unstructured interferences," *IEEE Trans. Signal Process.*, vol.51, no.2,pp.362-372,2003
- [15] E. G. Larsson, P. Stoica, and J. Li, "On maximum-likelihood detection and decoding for space-time coding systems," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 937-944, 2002.
- [16] L.Zheng and D. Tse, "Communicating on the Grassmann Manifold: A Geometric Approach to the Non-coherent Multiple Antenna Channel", *IEEE Transactions on Information Theory*, vol. 48(2), pp. 359-383, February 2002.